Data Management Requirements for Intensity Frontier Experiments (DRAFT)

CD/REX Department

v0.1

9 Nov 2010

Introduction

We discuss here the Data Management requirements for the IF experiments. This includes file systems used by these projects, and the associated meta-data, planning and management tools.

The issue of file deliver to batch and Grid processes is discussed separately in the Data Handling Requirements document.

Storage Elements

We give a brief outline of storage systems presently in use or investigation.

Local files

Interactive and local worker nodes may have of order 200 GBytes per core of local disk storage.

Grid works nodes have less, under 50 GBytes per core.

AFS

Use of AFS is primarily historical. Minos has moved its software releases and data handling to Bluearc. User login areas are still in AFS for most IF projects, but this could change.

AFS is still the primary location for the CD Web Servers. Bluearc is also supported, and we are like to shift there soon.

NFS

The GPCF has about 20 TBytes of high performance scratch space, intended to replace local disk storage

Bluearc

The bulk of IF disk storage is in Bluearc. We have deployed 10 to 100 TBytes per experiment.

Bluearc is a proprietary high performance NFS server.

These systems are available throughout Fermilab, including Fermigrid computing elements.

This has become the primary working space for IF experiments.

Software is installed in 'app' areas, mounted readonly and executeable on Fermigrid nodes.

Data is stored in 'data' aread, mounted writeable and non-executable on Fermigrid nodes.

DCache

Data is generally written to the Enstore archives via DCache write pools.

Data is read from the archives via DCache read pools, presently about 15 TB Public and 35 TB Minos

Enstore/tape

All permanent archives are via the Fermilab Enstore system, backed primarily by LTO-4 tape.
We duplicate critical files (raw data and important archives) in physically separate locations.

Investigations

We are participating in LCG storage studies, and the Fermilab Grid Storage Investigation.

Lustre - is in production use by the HPC group and is being studied by CMS.

Hadoop - is being investigated.

Meta-data

File system

The great majority of file transactions are by individual analysis users. They work with what is provided by the file system : file names, paths and sizes.

SAM

Minos, Minerva and Nova are using SAM metadata for officially managed files.

All SAM users (CDF, D0, Minos, Minerva, Nova) use the same back-end Oracle schema.

Typical file meta-data include file dates, sizes, checksums, parentage, data streams/tiers/types/famiies, and application names and versions.

Parameters can be added without making a schema change, but with caution. Each distinct parameter value creates a new row in the database. So parameters are good for things like a short list of strings, but not for floating point values. For example, Minos Monte Carlo data uses parameters to describe the Beam configuration, particle flavors, software release and vertex regions. Each of these have only a few possible values.

Luminosity

Historically, associating luminosity with files is a difficult task, handled differently by each experiment. There are basic hooks for this in SAM, but the data is handled differently by each experiment.

File listings

The Enstore system provides a daily complete file listing, needed for making global file scans.

Minos creates frequent file usage summaries for its AFS and Bluearc systems, to discourage users from doing full file scans.

Monitoring

Performance

Reliability and data transfer rates are being monitored for Bluearc. We should add this for the other systems.

Quotas

User quota are important for shared files systems written by individual users. This is both to prevent denial of service, and to track usage.

Summaries

It is important to track overall file system usage by user, project, and data category.

This is critical both for operations, and for capacity planning.